Evaluating Administrative Records as a Potential Sample Frame for the National Survey of College Graduates

by

John Voorheis
U.S. Census Bureau

CARRA Working Paper 18-14 **September, 2021**

Abstract

The National Survey of College Graduates (NSCG) is an important source of information on the education and career paths of the college graduate population in the United States. The NSCG currently uses the American Community Survey as a frame for sampling new respondents. There is interest, however, in whether other, alternative datasets may be superior sources for sample frame construction. In this paper, we examine the suitability of one of these datasets – the National Student Clearinghouse – by comparing its coverage and alignment with the ACS (the current frame source). Despite some deficiencies, the NSC has excellent coverage compared to the ACS, making it a candidate for potential supplementation, but not outright replacement of the current sampling frame.

Keyword*:* administrative records, college graduates, sampling frames

JEL Classification: C83, J24, C8

[*] This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed, DRB approval number CBDRB-FY18-408.

EXECUTIVE SUMMARY

The National Study of College Graduates (NSCG), sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF), provides important information about the education and career paths of the Nation's college graduates for policy analysis and general research. Academic researchers and government policymakers use these data to study the relationship between education and career opportunities.

Working with NCSES, the Demographic Research Group in the Center for Economic Studies (CES) at the U.S. Census Bureau (formerly the Center for Administrative Records Research and Applications) is researching ways to make use of administrative records in NSCG survey operations. This report describes one part of this task – investigating alternate data sources which could supplement or replace the current NSCG sample frame, which is drawn from the American Community Survey (ACS). The analysis in this report provides some preliminary evidence for whether one such alternate data source – the National Student Clearinghouse (NSC) – may be suitable for frame enhancement or replacement.

CES has access to a subset of the NSC data that contains college enrollment information for all graduates of public high schools in the state of Oregon. While this subset is limited along geographic and temporal lines, it nonetheless provides an opportunity for an investigation of the data quality and coverage of NSC records. Using this subset of the full NSC data, we analyze the degree to which the NSC's coverage overlaps with the current source for the NSCG sampling frame – the ACS. Additionally, we investigate the degree to which ACS responses to questions on enrollment, graduation and field of degree align with data in the NSC.

When properly constrained to the universe of Oregon public high school students, we find that most in-universe ACS college graduates can be linked to the NSC data, suggesting a high degree of coverage relative to the current frame. There is some variability across demographic groups and occupational categories which suggests that the coverage of the NSC data relative to the ACS may be worse for disadvantaged or low socio-economic status (SES) individuals. Additionally, across demographic breakouts, coverage is worse for 2-year college graduates than for 4-year college graduates, which highlights a potential shortcoming of the full NSC data.

When ACS respondents can be linked to the NSC data, there is a high degree of agreement between ACS responses to enrollment, educational attainment and field of degree questions. About 82 percent of linked cases exhibit agreement between the ACS and NSC on enrollment, and about 95 percent agree on graduation. A majority of major field of degree groups exhibit close to 100 percent agreement between the two sources. However, there is slightly worse agreement conditional on NSC graduation for 2-year degree holders – only about 53 percent of individuals with an NSC graduation record report having a 2-year degree in the ACS.

Overall, the high rate of coverage for most groups and the high degree of agreement on 4-year college graduation point towards the NSC being a potentially useful source of data for

supplementing the existing frame. The poorer coverage of 2-year degree holders, and poor agreement rates for graduation from 2-year institutions, however, are signals against suitability. This, combined with the relatively lower rate of linkage to graduation records than to any enrollment record and the lack of coverage for older college graduates suggest that the NSC is not a suitable replacement for the current frame. Future work on whether the NSC can be used to improve the current ACS based frame is necessary to better understand how the variation in coverage for certain groups can be mitigated. Such future work may take advantage of rich residential history information combined with demographic information to identify households which likely contain college graduates who otherwise would not be sampled in the NSCG. Separate from issues around frame coverage, these results do suggest that, were NSC data to be made more broadly available in survey production, improvements to editing, imputation and weighting may be possible.

Contents

Background

The National Survey of College Graduates is a biennial survey of college graduates residing in the United States sponsored by NCSES within NSF. The U.S. Census Bureau serves as the NSCG data collection contractor on behalf of NCSES. The Center for Economic Studies at the US Census Bureau has been tasked by NCSES with investigating whether administrative records resources currently held as part of the Census Bureau's data infrastructure can be used to improve survey measurement and operations for the NSCG. This report describes one part of this investigation, focusing on potential administrative records sources to augment or replace the current NCSG sampling frame.

Currently, the NSCG sampling frame is drawn from addresses on the Census Master Address File (MAF) which have been previously sampled in the American Community Survey (ACS), in which at least one member of the household reported having a college degree. This has the advantage of being a very efficient frame -- costs due to a large number of screeners being returned by non-college graduate households are reduced. However, because it is based on the ACS, any sampling error in the ACS will propagate into the NSCG. Thus, an alternate administrative records approach to constructing the NSCG sample frame is attractive, as it may improve coverage, reduce sampling error and/or reduce costs.

To construct an alternate administrative records based frame, two components are required: an administrative records source on college graduates or individuals who have attended college, and an administrative records source linking individuals to their locations. CES has access to several administrative records datasets which can be employed as the second component, and there is ongoing work to combine these sources to create a harmonized residential history file which could be used for survey improvement work. CES has identified several potential datasets which could be employed for the former component. These potential datasets include a) form 1098-T filings from the IRS, b) the National Student Loan Data System (NSLDS) and c) data from the National Student Clearinghouse (NSC). Unfortunately, CES does not have access to universe files for any of these data sources, but does have an extract of the NSC data for the universe of students enrolled in Oregon public schools between 2004-2015.

Thus the NSC will be the focus of the analysis in this report. This is due not just to short term feasibility, however. The NSC may provide a superior source for frame information compared to 1098-T or NSLDS data. A vast majority of higher education institutions participate in the NSC, providing student records for all enrolled students on a quarterly or semester basis. On the other hand, the NSLDS only covers individuals who received federal student loans, and the IRS only requires universities to file 1098-T for students who paid qualified educational expenses in a

given year. Additionally, there are medium term barriers to data access for survey improvement for all three datasets, especially for the NSLDS.[1]

Data, Methodology and Research Questions

Currently, the NSCG uses the ACS as a sample frame. To assess the utility of the NSC as a potential alternate frame, we will report the degree to which ACS respondents can be located in the NSC. That is to say, we are interested in how well the proposed alternate frame covers the existing frame. We are particularly interested in two sub-populations of ACS respondents: individuals who are currently enrolled in a post-secondary institution, and individuals who have a post-secondary degree. Specifically, we will seek to answer the following research questions:

1) How many in-universe ACS respondents with college degrees or who are currently attending college can be matched to NSC records?
    a. Does this vary by sex, race, Hispanic origin, occupation or field of degree?
2) Among in-universe ACS individuals who can be matched to NSC records, to what degree are ACS and NSC information on enrollment, graduation and field of degree in agreement?

To answer these questions, we will link ACS microdata with an extract of NSC data held by CES using unique anonymous identifiers called Protected Identification Keys (PIKs), which are assigned to microdata files using a probabilistic matching algorithm called the Person Validation System (PVS). This extract was acquired as part of a joint statistical project with the Oregon Department of Education and outside researchers. As part of the joint statistical project, CES obtained administrative records containing personally identifiable information (PII) for all Oregon public school students who attended school between 2004-2015. CES staff then assigned PIKs to these records, and then requested an extract of NSC records for all SSNs associated with the unique PIKs. These unique set of PIKs will serve as the universe for the purposes of our analysis – these are the set of all linkable individuals who can potentially be found in both datasets.

We will focus our analysis on this universe for the coverage and agreement analysis, as it is the cleanest way of doing this. It is, however, also possible to do coverage analysis based on three approximations of the administrative records universe using ACS and other administrative records. These approximations include a) ACS respondents residing in Oregon, b) ACS respondents who were born in Oregon in the Numident[2], and c) ACS respondents who lived in Oregon between the ages of 14-18 according to the Master Address File Auxiliary Reference File

[1] Resolving the privacy language in FERPA (which allows the use of PII for individual students only for statistical purposes which benefit Department of Education programs) and Title 13 (which allows the use of PII for individuals on for statistical purposes which benefit Census Bureau programs) is a difficult and thus far intractable task.
[2] The Social Security Administration's Numerical Identification File (Numident) is master file of all Social Security Numbers which have been issued. The Census Bureau receives an extract of this database which serves as the master file for the PVS process.

(MAF-ARF), a composite person-place administrative records file. We present additional results using these approximations to the universe in the appendix.

It should be noted that the National Student Clearinghouse data is collected from voluntary submissions by post-secondary institutions, and thus does not have complete coverage of the college attending or college degree holding populations. The number of post-secondary institutions covered by NSC has grown over time, and a vast majority (80 percent or more) of currently enrolled students are covered after about 2000. There is, however, some well-known heterogeneity in this coverage, which has evolved over time. In particular, coverage in 2-year institutions is lower than in 4-year institutions, and in private for-profit institutions than in private non-profit or public institutions.[3] These gaps have narrowed somewhat in recent years -- the gap between public 2-year and 4-year institutions has largely been erased by 2017 – however for our period of analysis gaps in coverage are meaningful. Additionally, although institutions are able to report graduation to the NSC, there is some concern that this is not universally reported, especially in the case of certificates awarded by 2-year institutions (Cook and Pullaro, 2010). We will perform a series of subgroup analyses to examine the degree to which these gaps in coverage are potentially driving overall match rates.

Our analysis proceeds as follows. We pool all household and group quarters observations from the 2009-2015 ACS with valid PIKs. We then restrict this sample to the set of people who are between the ages of 18-30 at ACS response and who would have been 18 or younger in 2004. We then match these observations by PIK to the NSC extract described previously, retaining only observations which are in the universe of valid PIKs for the Oregon Department of Education administrative data.

Results

ACS Coverage

We begin by examining how the CES NSC Extract compares in coverage to the sub-sample of linkable ACS responses which can be linked to Oregon public school admin data. Note that because this analysis is constrained to a specific extract, these results may not apply to coverage of the entire NSC. This comparison will be presented as the proportion of in-universe individuals who appear in the NSC extract data. We perform this comparison for several sub-groups in light of known patterns of coverage and potential under-reporting by institutions outlined above. These subgroups for college attendance include: 1) all ACS college attendees, 2) ACS college attendees at 4-year institutions, and 3) ACS college attendees at 4-year institutions between the ages of 19-22. While for college graduates they include: 4) all ACS college graduates, 5) ACS 4-year college graduates, 6) All ACS college graduates listed as graduates in the NSC, and 7) ACS 4-year college graduates listed as college graduates in the NSC.

[3] For more information on coverage of the NSC, see Dynarski, et al. (2011) and https://nscresearchcenter.org/workingwithourdata/

Table 1 summarizes the overall coverage rates for these subgroups for the in-universe ACS subsample. Note that in this and subsequent tables, unweighted counts are rounded in accordance with Census disclosure avoidance rules, and all proportions are calculated from the rounded counts. Table 1 shows the unweighted count of in-universe ACS respondents, the unweighted count of in-universe ACS respondents who can be found in the NSC data, and the associated percentage calculation. Overall, coverage rates are quite high, especially relative to previously released results (see appendix). Coverage is highest for ACS college graduates from 4-year institutions, but much lower for current attendees – only 86.84 percent of ACS in-universe current undergraduate attendees can be found in the NSC data. Similarly, while ACS college graduates can be found in the NSC data, they do not always have a graduation record – only 76 percent of college graduates (and only 87.5 percent of graduates of 4-year institutions) can be matched to an NSC graduation record.

Table 1: Coverage of CES NSC Extract vs. ACS 2009-2015

Subgroup	ACS N	NSC N	PCT
Any College Attendance	5700	5000	87.72
UG College Attendance	5400	4700	87.04
UG College Attendance, 19-22	3800	3300	86.84
AA+ College Graduates	2500	2300	92
BS+ College Graduates	1600	1600	100
AA+ College Graduates (Grad in NSC)	2500	1900	76
BS+ College Graduates (Grad in NSC)	1600	1400	87.5

Source: ACS 2009-2015, CES NSC Extract.

Variation in coverage across the subgroups in Table 1 lines up with our knowledge about the overall quality of the NSC data – coverage rates are much higher for the 4-year degree college graduates than for the overall college graduate population, implying lower coverage of 2-year degrees. While the coverage rates are reasonably high, the subgroup results do suggest a few areas of concern, not limited to the fact that 2-year degrees seem to be under-covered. In particular, the pattern that degree holders in the ACS are less likely to have a graduation record in the NSC than they are to have an enrollment record in the NSC suggests a trade off. If the NSC were to be used as a frame source, sampling from individuals with graduation records would systematically miss a substantial fraction of cases (~12.5 percent of ACS college graduates), while sampling from all individuals with enrollment records would result in false positives – individuals who did not in fact graduate from college.

Although overall coverage rates are relatively high, this may not necessarily translate to high coverage rates for specific socio-demographic subgroups. Tables 2 and 3 break out the coverage calculations in Table 1 by Sex, Hispanic Origin and Race. Coverage rates are slightly higher for

women than for men, particularly when comparing ACS college graduates to NSC records with a graduation flag.

Table 2: Coverage of CES NSC Extract vs. ACS 2009-2015, by Sex and Hispanic Origin

Subgroup	ACS N	NSC N	PCT
Males			
Any College Attendance	2900	2500	86.21
UG College Attendance	2700	2400	88.89
UG College Attendance, 19-22	1900	1700	89.47
AA+ College Graduates	1100	1000	90.91
BS+ College Graduates	750	700	93.33
AA+ College Graduates (Grad in NSC)	1100	850	77.27
BS+ College Graduates (Grad in NSC)	750	600	80
Females			
Any College Attendance	3400	3000	88.24
UG College Attendance	3200	2900	90.63
UG College Attendance, 19-22	2300	2000	86.96
AA+ College Graduates	1600	1500	93.75
BS+ College Graduates	1100	1000	90.91
AA+ College Graduates (Grad in NSC)	1600	1300	81.25
BS+ College Graduates (Grad in NSC)	1100	950	86.36
Hispanic Origin			
Any College Attendance	450	350	77.78
UG College Attendance	450	350	77.78
UG College Attendance, 19-22	300	250	83.33
AA+ College Graduates	150	150	100
BS+ College Graduates	90	80	88.89
AA+ College Graduates (Grad in NSC)	150	100	66.67
BS+ College Graduates (Grad in NSC)	90	60	66.67
Non-Hispanic Origin			
Any College Attendance	5800	5200	89.66
UG College Attendance	5500	4900	89.09
UG College Attendance, 19-22	3900	3500	89.74
AA+ College Graduates	2600	2400	92.31
BS+ College Graduates	1700	1600	94.12
AA+ College Graduates (Grad in NSC)	2600	2000	76.92
BS+ College Graduates (Grad in NSC)	1700	1500	88.24

Source: ACS 2009-2015, CES NSC Extract.

The bottom two panels of table 2 report coverage broken out by Hispanic origin, while table 3 reports coverage broken out by race.[4] Coverage rates are higher for non-Hispanics than for Hispanics – only about 2/3rds of Hispanic origin ACS 4-year college graduates can be linked to NSC records with a graduation flag, compared to about 88 percent for non-Hispanics. Patterns of relative under- or over-coverage across race groups are less clear, however. Blacks and Asians seem to have slightly lower coverage rates than Whites, although some sub-groups are difficult to compare due to small sample sizes, resulting in cells being suppressed for disclosure avoidance.

Table 3: Coverage of CES NSC Extract vs. ACS 2009-2015, by Race

Subgroup	ACS N	NSC N	PCT
White			
Any College Attendance	5600	5000	89.29
UG College Attendance	5300	4700	88.68
UG College Attendance, 19-22	3700	3300	89.19
AA+ College Graduates	2500	2300	92
BS+ College Graduates	1700	1600	94.12
AA+ College Graduates (Grad in NSC)	2500	2000	80
BS+ College Graduates (Grad in NSC)	1700	1400	82.35
Black			
Any College Attendance	150	100	66.67
UG College Attendance	150	100	66.67
UG College Attendance, 19-22	90	80	88.89
AA+ College Graduates	40	30	75
BS+ College Graduates	20	20	100
AA+ College Graduates (Grad in NSC)	40	20	50
BS+ College Graduates (Grad in NSC)	20	(D)	(D)
Some other race			
Any College Attendance	100	100	100
UG College Attendance	100	100	100
UG College Attendance, 19-22	70	60	85.71
AA+ College Graduates	30	30	100
BS+ College Graduates	20	20	100
AA+ College Graduates (Grad in NSC)	30	20	66.67
BS+ College Graduates (Grad in NSC)	20	(D)	(D)
American Indian or Alaska Native			

[4] Throughout, we use the current two question format to assign race and ethnicity, treating Hispanic origin and race as two separate, non-mutually exclusive categorizations.

Any College Attendance	100	100	100
UG College Attendance	100	100	100
UG College Attendance, 19-22	70	60	85.71
AA+ College Graduates	40	30	75
BS+ College Graduates	20	20	100
AA+ College Graduates (Grad in NSC)	40	20	50
BS+ College Graduates (Grad in NSC)	20	20	100
Asian			
Any College Attendance	300	250	83.33
UG College Attendance	250	250	100
UG College Attendance, 19-22	200	150	75
AA+ College Graduates	100	100	100
BS+ College Graduates	90	80	88.89
AA+ College Graduates (Grad in NSC)	100	90	90
BS+ College Graduates (Grad in NSC)	90	70	77.78
Native Hawaiian or Pacific Islander			
Any College Attendance	30	20	66.67
UG College Attendance	30	20	66.67
UG College Attendance, 19-22	20	20	100
AA+ College Graduates	(D)	(D)	(D)
BS+ College Graduates	(D)	(D)	(D)
AA+ College Graduates (Grad in NSC)	(D)	(D)	(D)
BS+ College Graduates (Grad in NSC)	(D)	(D)	(D)

Source: ACS 2009-2015, CES NSC Extract. Note: (D) indicates a cell has been suppressed to avoid disclosure.

We break down the coverage analysis along two additional dimensions capturing potential populations of interest to NCSES – occupation groups and field of degree. Unlike the demographic breakdowns above, there are a large number of occupation and field of degree categories. Even the broadest (2 digit) occupation codes and field of degree codes in the ACS have dozens of unique categories. Many of these occupation groups and fields of degree are sparsely populated, given our relatively small sample. Thus, to avoid reporting multiple rows of suppressed cells, we collect only the five most common occupation groups and only report fields of degree where we can disclose unweighted counts.

Table 4 breaks down the coverage for college graduates by the five largest occupation groups (sales, management, office and administrative support, food preparation and education). NSC coverage is lower for sales and for food preparation related occupations than for management, office or education related occupations. Only 75 percent of ACS 4-year college graduates in sales occupations can be linked to NSC records, compared to approximately 100 percent for individuals in management and education occupations. The lower coverage occupation groups

roughly line up with lower wage, lower socioeconomic status professions relative to the high coverage occupations – mean annual wages for individuals in management is nearly three times as large as for those in sales (120,000 vs.40,700).[5]

Table 4: Coverage of CES NSC Extract vs. ACS 2009-2015, by Occupation

Subgroup	ACS N	NSC N	PCT
Sales and Related Occupations			
AA+ College Graduates	300	250	83.33
BS+ College Graduates	200	150	75
AA+ College Graduates (Grad in NSC)	300	200	66.67
BS+ College Graduates (Grad in NSC)	200	150	75
Office and Administrative Support Occupations			
AA+ College Graduates	400	400	100
BS+ College Graduates	250	250	100
AA+ College Graduates (Grad in NSC)	400	300	75
BS+ College Graduates (Grad in NSC)	250	250	100
Food Preparation and Serving Occupations			
AA+ College Graduates	250	200	80
BS+ College Graduates	100	90	90
AA+ College Graduates (Grad in NSC)	250	150	60
BS+ College Graduates (Grad in NSC)	100	80	80
Management			
AA+ College Graduates	150	150	100
BS+ College Graduates	100	100	100
AA+ College Graduates (Grad in NSC)	150	100	66.67
BS+ College Graduates (Grad in NSC)	100	100	100
Education, Training, and Library Occupations			
AA+ College Graduates	250	250	100
BS+ College Graduates	250	250	100
AA+ College Graduates (Grad in NSC)	250	250	100
BS+ College Graduates (Grad in NSC)	250	200	80

Source: ACS 2009-2015, CES NSC Extract.

Table 5 breaks down the coverage calculations by field of degree groups reported on the ACS for ACS 4-year college graduates. These field of degree groups are defined by the first two digits of the ACS field of degree codes (these correspond to the first two digits of the CIP codes in the NSC data). Given sample size constraints, there are a number of field of degree groups that are

[5] Wage data from the May 2017 National Occupational Employment and Wage Estimates United States, https://www.bls.gov/oes/current/oes_nat.htm

too sparsely populated to disclose results. There is less of a clear pattern of relative over- or under-coverage by college major than for the other demographic breakdowns. Due to rounding, we report most major groups as having coverage of approximately 100 percent. Notable exceptions include Communications majors, who have 89 percent coverage, Physical Sciences, who have 86 percent coverage and Social Science majors, who have 75 percent coverage.

Table 5: Coverage of CES NSC Extract vs. ACS 2009-2015, by Field of Degree

Major	Subgroup	ACS N	NSC N	PCT.
Business	BS+ College Graduates	300	300	100
Business	BS+ College Graduates (Grad in NSC)	300	250	83.33
Engineering	BS+ College Graduates	90	90	100
Engineering	BS+ College Graduates (Grad in NSC)	90	80	88.89
Education	BS+ College Graduates	100	100	100
Education	BS+ College Graduates (Grad in NSC)	100	90	90
Physical Sciences	BS+ College Graduates	70	60	85.71
Physical Sciences	BS+ College Graduates (Grad in NSC)	70	60	85.71
Family and Consumer Sciences	BS+ College Graduates	40	30	75
Family and Consumer Sciences	BS+ College Graduates (Grad in NSC)	40	30	75
Psychology	BS+ College Graduates	100	100	100
Psychology	BS+ College Graduates (Grad in NSC)	100	100	100
Social Sciences	BS+ College Graduates	200	150	75
Social Sciences	BS+ College Graduates (Grad in NSC)	200	150	75
Interdisciplinary Studies	BS+ College Graduates	20	20	100
Interdisciplinary Studies	BS+ College Graduates (Grad in NSC)	20	20	100
Biology and Life Sciences	BS+ College Graduates	100	100	100
Biology and Life Sciences	BS+ College Graduates (Grad in NSC)	100	100	100
Environment and Natural Resources	BS+ College Graduates	30	30	100
Environment and Natural Resources	BS+ College Graduates (Grad in NSC)	30	30	100
Medical and Health Sciences	BS+ College Graduates	100	100	100
Medical and Health Sciences	BS+ College Graduates (Grad in NSC)	100	100	100
Fine Arts	BS+ College Graduates	100	100	100
Fine Arts	BS+ College Graduates (Grad in NSC)	100	90	90
Communications	BS+ College Graduates	90	80	88.89
Communications	BS+ College Graduates (Grad in NSC)	90	70	77.78
Agriculture	BS+ College Graduates	30	30	100
Agriculture	BS+ College Graduates (Grad in NSC)	30	20	66.67
History	BS+ College Graduates	50	50	100
History	BS+ College Graduates (Grad in NSC)	50	40	80
Liberal Arts and Humanities	BS+ College Graduates	20	20	100
Liberal Arts and Humanities	BS+ College Graduates (Grad in NSC)	20	20	100
Computer and Information Sciences	BS+ College Graduates	30	30	100
Computer and Information Sciences	BS+ College Graduates (Grad in NSC)	30	20	66.67

Physical Fitness	BS+ College Graduates	50	50	100
Physical Fitness	BS+ College Graduates (Grad in NSC)	50	50	100
English	BS+ College Graduates	70	70	100
English	BS+ College Graduates (Grad in NSC)	70	60	85.71
Linguistics and Foreign Languages	BS+ College Graduates	50	50	100
Linguistics and Foreign Languages	BS+ College Graduates (Grad in NSC)	50	40	80
Mathematics and Statistics	BS+ College Graduates	20	20	100
Mathematics and Statistics	BS+ College Graduates (Grad in NSC)	20	20	100
Philosophy and Religious Studies	BS+ College Graduates	20	20	100
Philosophy and Religious Studies	BS+ College Graduates (Grad in NSC)	20	20	100

Source: ACS 2009-2015, CES NSC Extract.

Overall, coverage of the NSC extract relative to the current source data for the NSCG is relatively high, at 92 percent overall for all 2 and 4 year ACS college graduates. There are notable caveats to this overall high level of coverage – chiefly, fewer ACS college graduates can be matched to NSC graduation records than NSC enrollment records. This high level of coverage is not uniform across different subgroups. Coverage is lower for Hispanics, non-whites and individuals in lower-paid occupations; coverage is also lower overall for individuals with 2-year degrees. Although the high degree of coverage relative to the current frame source is heartening, the lower coverage for groups of potential interest should give us pause.

ACS-NSC Agreement

Although the primary potential use for the NSC data is likely to be as an additional or replacement data source for the NSCG frame, the NSC data can also inform measurement error by allowing the comparison of survey responses to equivalent concept administrative records data elements. We focus on the ACS questions that can be compared to NSC data – college enrollment, college graduation and field of degree. Validating the former can yield information on the current source data for the NSCG frame – i.e., are there currently enrolled college students or college graduates missed in the ACS due to item non-response or misreporting? On the other hand, validating the latter provides a template for the use of NSC data to potentially supplement or replace NSCG survey questions on field of degree.[6]

For the following summaries of the agreement between ACS survey responses and equivalent information in the NSC data, we restrict our attention to the set of ACS respondents who can be linked to the NSC data at least once. Table 7 summarizes the agreement between ACS responses to the college enrollment question and NSC enrollment information. Since NSC records are reported for each school year, we define a case as being in agreement if there is an NSC

[6] It would in principle be possible to validate field of degree information in the NSCG with the full NSC data, however using the CES NSC extract matched to the NSCG , would result in a prohibitively small sample relative to the ACS data linked to the NSC.

enrollment record corresponding to the month and year of ACS response, and disagreement for all cases where an enrollment record cannot be found, conditional on being in the NSC data at least once. Overall, NSC and ACS enrollment match reasonably well – about 82 percent of cases agree between the two sources. However, of the cases where the sources disagree, a majority occur where the NSC data says the student was enrolled but the student did not report attending college.

Table 7: CES NSC Extract vs. ACS 2009-2015 Agreement, Current Enrollment

	Not Enrolled in College in NSC	Enrolled in College in NSC
Not Enrolled in College in ACS	10500	1800
Enrolled in College in ACS	1600	5200
Overall Agreement:		82.2

Source: ACS 2009-2015, CES NSC Extract.

Table 8 repeats the agreement exercise for the analysis of agreement between ACS and NSC data on graduation. Since both the ACS and NSC contain information about the type of institution (2 vs. 4 year) we breakdown agreement rates for these types of institutions separately, with 2-year graduation agreement rates in the top panel and 4-year graduation agreement rates in the bottom panel. Agreement is very high, and much higher than for enrollment – about 95 percent of cases agree for 2-year graduation, compared to about 96 percent for 4 year graduation. These high agreement rates are largely a product of the high agreement between non-graduates. The marginal match rates tell a substantially different story – of individuals who report graduating from a 4-year college in the ACS, only 79 percent have a graduation record in the NSC; this is 50 percent for 2-year schools. Similarly, of linked cases where an NSC graduation record is present, about 83 percent report being graduates of 44-year institutions (53 percent for 2-year institutions). This suggests that there may be both substantial false positive error and false negative error, which may adversely affect the quality of NSCG data. There are both households improperly included in sample (who report graduation in the ACS but are not NSC graduates) and households improperly excluded from the sample (who report not being graduates in the ACS but in fact did graduate according to NSC).

Table 8: CES NSC Extract vs. ACS 2009-2015 Agreement, Graduation

2 Year College Graduation

	Not College Graduate in NSC	College Graduate in NSC
Not College Graduate in ACS	18000	450
College Graduate in ACS	500	500
Overall Agreement:		95.12

4 Year College Graduation

	Not College Graduate in NSC	College Graduate in NSC
Not College Graduate in ACS	17000	350
College Graduate in ACS	450	1700
Overall Agreement:		95.9

Source: ACS 2009-2015, CES NSC Extract.

Table 9 reports agreement rates between the field of degree reported by ACS respondents and the field of degree reported in the NSC. Note that since the NSC field of degree element is only non-missing for records with a graduation flag, so our universe for this table is 4-year college graduates in the ACS who can be linked to an NSC record with a graduation flag. There is a wide variation in the degree of agreement between the ACS and NSC. The lowest agreement (as low as 50 percent) is for Physical Science, Education and Environmental and Natural Resources related majors. Majors such as Biology, Mathematics and Social Science agree in almost all cases.

Table 9: CES NSC Extract vs. ACS 2009-2015 Agreement, Field of Degree

Major	# of ACS Cases	# of ACS Cases in Agreement with NSC	Agreement Rate
Agriculture	30	30	100
Biology and Life Sciences	150	150	100
Business	300	250	83.33
Communications	100	70	70
Computer and Information Sciences	40	30	75
Education Administration and Teaching	100	60	60

Engineering	100	90	90
English Language, Literature, and Composition	60	60	100
Environment and Natural Resources	30	20	66.67
Family and Consumer Sciences	40	30	75
Fine Arts	100	80	80
History	50	40	80
Liberal Arts and Humanities	30	20	66.67
Linguistics and Foreign Languages	50	40	80
Mathematics and Statistics	20	20	100
Medical and Health Sciences and Services	100	100	100
Physical Fitness, Parks, Recreation, and Leisure	50	50	100
Physical Sciences	80	40	50
Psychology	100	90	90
Social Sciences	200	200	100

Source: ACS 2009-2015, CES NSC Extract.

Conclusion and Future Work

Using the Census Bureau's data linkage infrastructure, we have linked responses to the ACS (the current source data for the NSCG frame) with an extract from the National Student Clearinghouse (a potential source for supplementing or replacing the current frame). By analyzing the coverage, data quality, and alignment of the NSC extract data with the ACS, we provide important preliminary evidence for the suitability of the NSC as an alternate source of data for the NSCG frame.

Overall, when restricted to the relevant universe (Oregon public high school graduates), the coverage of the NSC extract held by CES relative to the ACS is quite high – over 90 percent of ACS college graduates can be linked to the NSC data. Additionally, overall there is a high degree of agreement between ACS responses and NSC data on questions on enrollment, graduation and field of degree. About 95 percent of linked cases exhibit agreement between ACS and NSC on graduation, and about 82 percent of cases agree on enrollment.

However, there are some subgroups for whom the overall high rate of coverage and agreement do not hold up. In particular, coverage is lower for graduates from 2-year institutions, and for individuals who are from disadvantaged communities or who may be low SES. There is also lower agreement between the ACS and NSC on graduation for 2-year degree holders – only 53 percent of in-universe individuals who have an NSC graduation record report having a 2-year college degree on the ACS. Together, these results, which are consistent with known deficiencies in the NSC data collection process, suggest that the NSC may not be a suitable stand-alone frame data source for this population of interest.

Although the NSC may not be suitable for a full replacement for the current frame, there is reason to believe that it could be useful as a data source for supplementing or otherwise improving the frame. Future work on the feasibility of this supplementation will require data access which the Census Bureau

does not currently have, however. In order to proceed, the Census would need to acquire a much larger subset (or the full universe) of NSC cases under a new agreement with the data owners. However, if such an agreement can be concluded, there are several possibilities for integrating the NSC into the NSCG operations.

One possibility would be to design a hybrid frame which samples from both ACS responding households as well as households which can be linked to an NSC record. We observe, for instance, that a non-trivial number of ACS respondents who can be linked to an NSC record do not report attending or graduating college. One approach, then, would be to sample not just from not just ACS respondents who report having a college degree, but also ACS respondents who can be linked to the NSC who do not report receiving a college degree – this approach could address some of the under-coverage of college-educated individuals reported on the ACS and, as a result, improve the NSCG's coverage. An additional approach would be to use the NSC, combined with residential history information, to identify households which likely contain a college graduate but are not in the status quo ACS-based frame, adding a second stratum to the sampling process alongside the current ACS sampling frame. This approach would be similar to other frame work done by CES-Demo for the National Survey of Children's Health.

Additionally, since the NSC data contains some useful information on fields of degree and institution type, it may be possible to leverage the NSC to improve the quality of NSCG data after collection. For instance, it may be possible to use NSC-derived information in the edit and imputation process, either as an input into a model, or as a direct replacement for item non-response. If data use agreements can be satisfactorily concluded, linking previous waves of NSCG to the NSC would be useful to assess the degree to which these data sources agree, and the degree to which NSC data could be useful for improving NSCG data.

References

Cook, Bryan and Natalie Pullaro. “College Graduation Rates: Behind the numbers.” Working Paper, American Council on Education.

Dynarski SM, Hemelt SW, Hyman JM. The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes. *Educational Evaluation and Policy Analysis*. 2015;37:53-79.

Wagner, Deborah and Mary Layne. 2014. “The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications’ (CARRA) Record Linkage Software”. CARRA Working Paper no. 2014-01. U.S. Census Bureau, Washington, D.C.

Appendix 1: Alternative Coverage Comparisons

IPEDS Coverage

Table A1: Comparison of Aggregate Enrollment, NSC vs. IPEDS

School Year	IPEDS OR Enrollment	IPEDS Fall OR Enrollment	# Students Enrolled in OR Colleges in NSC	# Students Enrolled in OR Colleges in NSC (Fall)
2004-05	301,872	204,488	13,000	4,500
2005-06	309,837	203,262	32,000	20,500
2006-07	312,183	200,856	49,000	34,500
2007-08	324,469	206,164	66,000	47,500
2008-09	347,953	223,772	83,500	62,000
2009-10	368,965	246,507	99,500	73,000
2010-11	380,583	253,909	112,000	81,500
2011-12	382,427	263,019	119,000	87,500
2012-13	375,581	258,357	121,000	89,000
2013-14	366,391	254,675	113,000	84,000
2014-15	357,382	249,023	107,000	82,000

Source: ED IPEDS, CES NSC Extract

Comparing NSC to Approximations of OR Public School Universe

Table A2: Coverage of CES NSC Extract vs. ACS 2009-2015, Alternate Approximations to OR Universe

	In OR in ACS			Born in OR			Lived In Oregon at Age 5-18		
Subgroup	ACS N	NSC N	PCT	ACS N	NSC N	PCT	ACS N	NSC N	PCT
Any College Attendance	6200	4600	74.19	10000	5600	56	8200	5600	68.29
UG College Attendance	5900	4400	74.58	9300	5300	56.99	7600	5300	69.74
UG College Attendance, 19-22	3800	3100	81.58	6000	3700	61.67	4900	3700	75.51
AA+ College Graduates	3400	1900	55.88	6000	2500	41.67	4900	2500	51.02
BS+ College Graduates	2300	1200	52.17	4200	1700	40.48	3500	1700	48.57
AA+ College Graduates, 18-30 (Grad in NSC)	3400	1500	44.12	6000	2100	35	4900	2100	42.86
BS+ College Graduates, 18-30 (Grad in NSC)	2300	1100	47.83	4200	1600	38.1	3500	1600	45.71

Source: ACS 2009-2015, CES NSC Extract.